A01

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027137 A2

(51) International Patent Classification[7]: C07K 1/107, 1/04, 14/815, 14/705, A61K 47/48

(21) International Application Number: PCT/GB02/04352

(22) International Filing Date:
26 September 2002 (26.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0123262.8 27 September 2001 (27.09.2001) GB

(71) Applicant *(for all designated States except US)*: **ADPROTECH LIMITED** [GB/GB]; Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SMITH, Richard, Andrew, Godwin** [GB/GB]; AdProTech Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **BETLEY, Jason, Richard** [GB/GB]; AdProTech Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **ESSER, Dirk** [GB/GB]; AdProTech Limited, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(74) Agent: **DAVIES, Jonathan, Mark**; Reddie & Grose, 16 Theobalds Road, London WC1X 8PL (GB).

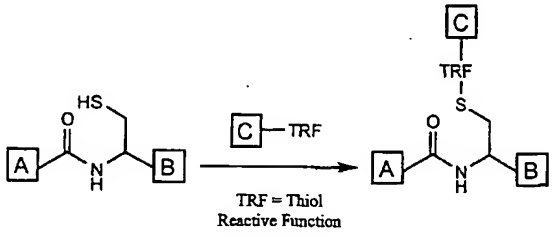
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

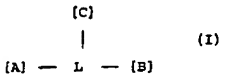
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYMERIC COMPOUNDS





(57) Abstract: The invention provides a process for forming a compound having the formula (I): from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] had a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] had a thiol reactive function (TRF) group, wherein chemical entities [A], [B] and [C] are covalently linked by linker group L, comprising the steps of: (i) admixing native [A] and native [B] in a reaction solution; (ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond; (iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group; (iv) admixing native [C] with the second intermediate compound in a reaction solution; and (v) reacting the thiol reactive function (TRF) group of native [C] with the free thiol group of the second intermediate compound for producing a compound having the formula (I).

WO 03/027137 A2

Polymeric Compounds

This invention relates to assembly of polymeric compounds using native chemical ligation as a method step. In particular, it relates to polypeptide assemblies, their use in therapy, and methods and intermediates for their production.

Background to the Field of Invention

Combination therapy is a widely used approach to many diseases in the clinic. It is found frequently that the effects of administering two or more drugs for the same indication of only modest efficacy may result in a clinical effect significantly greater than the sum of the parts. Such therapies are used widely in diseases such as cancer and AIDS.

Ligand directed therapy is another recent development that has received wide attention. The therapy relies on a pharmacophore (drug) being covalently or non-covalently attached to a ligand specific for a certain cell surface marker or tissue. The ligand directs the drug to the surface of the cell where it may exhibit its action at an increased local concentration. This may result in a lower dose being required, may reduce side-effects associated with systemic drug delivery, and thus may increase the therapeutic ratio. The drug-ligand conjugate may alternatively be taken up into the cell where it may exhibit its action, either in its conjugate form, or as free drug if a linker labile to intracellular conditions is incorporated into the conjugate. For instance, the linker may be labile to the acid conditions present in lysosomal compartments. The directing ligand may be a small molecule, peptide, binding protein, antibody, or antibody fragment, identified by one of a number of combinatorial binding screens, for example.

A few examples in the prior art provide for bifunctional drug molecules that combine both elements of a combination therapy regime covalently linked in the same molecule (see for example US5780653).

Linker molecules are commercially available with three homofunctional groups arrayed around a central core (e.g. TMEA - Pierce Chemical Co - Rockford, IL) that theoretically allow for such an approach. In practice, however, it is difficult to direct the absolute order of addition of the three components, and complex and lengthy purification steps may be required between the addition of each ligand or pharmacophore.

Another example of the same general principle is the starburst dendrimer system whereby successive rounds of addition to a central homopolyfunctional core results in a large number of reactive groups on the surface of a spherical macromolecule that is available for ligand/drug attachment (see EP0271180). Again, it is difficult to specifically control the ratios of ligand/drug loading with so many possible sites of attachment. In addition, the resultant derivatised dendrimer is extremely large and difficult to quantify exactly.

EP0832096 and Dawson, P.E. *et al.* (1994; Science 266: 776-799) disclose the ligation of two peptidic compounds to form proteins by a process known as native chemical ligation. The basic requirements for this coupling are the presence on participant A of a thioester moiety such as a benzyl or ethyl thioester, and the presence on participant B of an N-terminal cysteine or functional equivalent containing the minimal grouping of a 1-amino-2 thiol. The coupling reaction commences when the thio-nucleophile initially attacks the thioester in a reversible fashion causing the thioalkyl group to leave. The thioester so

formed then undergoes an intramolecular S→N acyl shift to irreversibly produce the ligated amide product **A-B** with a native cysteine (ie. as found naturally) with a free thiol group at the junction (Figure 1).

This process can be repeated if it is possible to unmask a further N-terminal cysteine at the N-terminal of **A**, in the product **A-B**, by enzymatic means or by classical protecting group chemistry. A further native chemical ligation coupling with a thioester-containing moiety **E** would therefore yield the product **E-A-B**. These methods have been used to generate a variety of proteins.

It is also possible to produce cyclic peptides/proteins by incorporating both the thioester and amino terminal cysteine moieties within the same molecule.

It is possible to perform the ligations described above without removing all of the reaction participants from a solid phase.

There is also a prior art method for incorporating a thioester moiety into a participant where one does not exist. This method relies on the reaction of a primary amine with thiolane-2,5-dione and subsequent alkylation of the thioacid formed with an appropriate alkyl bromide. This method does require a single primary amine to be present otherwise multiple thioesters may be coupled.

After a native chemical ligation (NCL) reaction has taken place, there exists at the junction site between **A** and **B** (Figure 1), a nascent cysteine sidechain (ie. a newly formed cysteine sidechain). If neither participant in the NCL reaction contained

a free thiol, the free thiol so formed will be the only free thiol in the molecule. No teachings in the prior art recognise the potential for using the free thiol described to react further with a thiol-reactive entity.

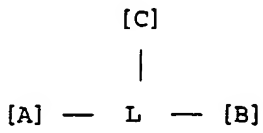
There are reports in which considerable effort has been made to remove the cysteine free thiol subsequent to native chemical ligation (WO 98/28434; Yan, L. Z. *et al.*, Journal of the American Chemical Society 123/4: 526-533; Botti, P. *et al.*, 2001, Tetrahedron Letters 42: 1831-1833; Low, D.W. *et al.*, 2001, Proceedings of the National Academy of Sciences of the USA 98[12]:6554-6559).

Detail of the Invention

In the present invention we show that the free thiol formed by native chemical ligation at the junction of **A** and **B** described earlier can in fact be derivatised with suitable reagents to form a branching point. A participant **C** may form a covalently-linked 'trimeric' entity comprising **A**, **B**, and **C** (for example: Figure 2).

The invention relates to the use of such assemblies in combinatorial ligand discovery. The invention teaches methods of assembling, for example, homo- and heterotrimers and tetramers by native chemical ligation/orthogonal grafting techniques. The invention is exemplified by a number of entities with therapeutic potential, where three or more functionalities are required to be arrayed in a predictable spatial fashion with defined molecular architecture at their junctions. Also described are a number of new classes of molecule that facilitate the assembly of molecular entities that would otherwise be unable to participate.

According to the present invention there is provided a process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

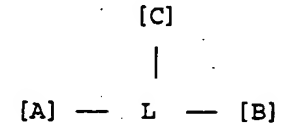
(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

(iv) admixing native [C] with the second intermediate compound in a reaction solution; and

(v) reacting the thiol reactive function (TRF) group of native [C] with the free thiol group of the second intermediate compound for producing a compound having the formula (I).

The terms "native A", "native B" and "native C" refer to the participant chemical entities in their form before reaction with each other. The "native" entities may thus be regarded as the pre-reaction entities.

Also provided is a process for the synthesis of a compound having the formula (I):



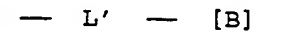
(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) reacting the thioester group of native [A] to the 1-amino-2-thiol group of native [B] by native chemical ligation to form an intermediate compound having the formula (II):



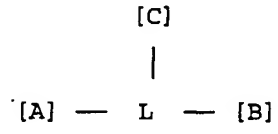
(II)

where intermediate linker group L' comprises an amide bond having attached thereto a free thiol;

and

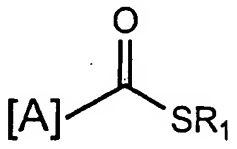
(ii) reacting the thiol reactive function (TRF) group of native [C] to the free thiol of the intermediate compound having the formula (II) to form the compound having the formula (I).

Yet further provided is a process for the synthesis of a compound having the formula (I):



(I)

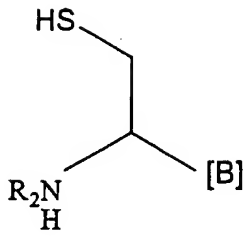
from chemical entities native [A], native [B] and native [C], where native [A] comprises a functional group having a chemical structure, excluding [A], shown in the formula (III):



(III)

where R_1 is H or a sidechain,

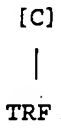
native [B] comprises a functional group having a chemical structure, excluding [B], shown in the formula (IV):



(IV)

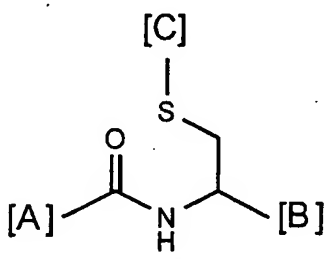
where R_2 is H or a sidechain,

and native [C] comprises a thiol reactive function (TRF) group attached to [C] as shown in formula (V):



(V)

wherein [A], [B] and [C] are chemical entities covalently linked by a linker group L, and wherein the compound of formula (I) has the chemical structure shown in formula (VI):



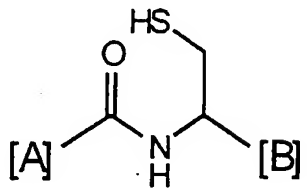
(VI)

comprising the steps of:

(i) forming an intermediate compound having the formula (II):

[A] — L′ — [B] (II)

wherein intermediate linker group L' comprises a free thiol group, and wherein the intermediate compound of formula (II) has the chemical structure shown in formula (VII):



(VII)

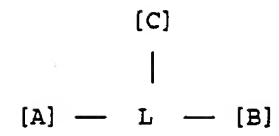
by reacting the functional group shown in formula (III) of native [A] to the functional group shown in formula (IV) of native [B] using native chemical ligation;

and

(ii) producing the compound having the formula (I) by reacting the thiol reactive function (TRF) group of native [C] to the free thiol of the intermediate linker group L'.

In the above-mentioned process of the invention, native [C] has a thiol reactive function (TRF) group for reacting with the thiol at the junction site between [A] and [B]. An alternative is to provide a thiol reactive group (TRF) linker or spacer linker which reacts with the thiol at the junction site between [A] and [B], then reacting the resultant molecule with native [C].

Thus further provided according to the present invention is a process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

(iv) admixing the second intermediate compound with a thiol reactive function (TRF) group linker having at least two thiol reactive function (TRF) groups in a reaction solution;

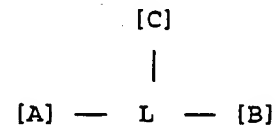
(v) allowing the thiol reactive function (TRF) group linker to react with the free thiol group of the second intermediate

compound to form a third intermediate compound with a thiol reactive function (TRF) group;

(vi) admixing native [C] with the third intermediate compound in a reaction solution; and

(vii) reacting the thiol group of native [C] with the thiol reactive function (TRF) group of the third intermediate compound for producing a compound having the formula (I).

Also provided is a process for the synthesis of a compound having the formula (I):



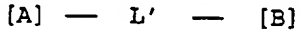
(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) reacting the thioester group of native [A] to the 1-amino-2-thiol group of native [B] by native chemical ligation to form a first intermediate compound having the formula (II):



(II)

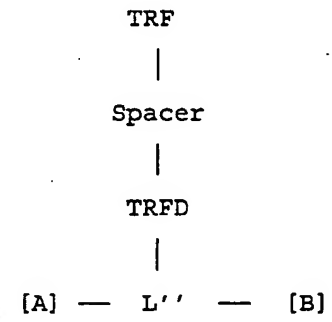
where intermediate linker group L' comprises an amide bond having attached thereto a free thiol group;

(ii) reacting a spacer linker having the formula (VIII):

TRF-Spacer-TRF (VIII)

wherein TRF is a thiol reactive function (TRF) group and the Spacer is a linking group,

with the free thiol of intermediate linker group L' to form a second intermediate compound having the formula (IX):



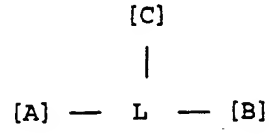
(IX)

where TRFD is a thiol reactive function derivative attached to second intermediate linker group L'',

and

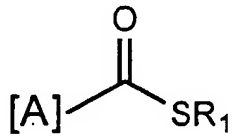
(iii) reacting the thiol of native [C] to the thiol reactive function (TRF) group of the second intermediate compound having the formula (IX) to form the compound having the formula (I).

Further provided is a process for the synthesis of a compound having the formula (I):



(I)

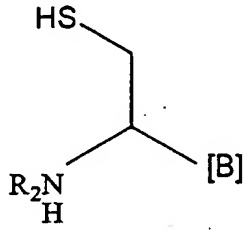
from chemical entities native [A], native [B] and native [C], where native [A] comprises a functional group having a chemical structure, excluding [A], shown in the formula (III):



(III)

where R_1 is H or a sidechain,

native [B] comprises a functional group having a chemical structure, excluding [B], shown in the formula (IV):



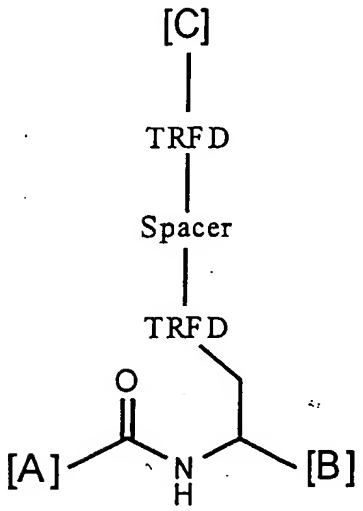
(IV)

where R_2 is H or a sidechain,

and native [C] comprises a thiol attached to [C] as shown in formula (X):

(X)

wherein [A], [B] and [C] are chemical entities covalently linked by a linker group L, and wherein the compound of formula (I) has the chemical structure shown in formula (XI):



(XI)

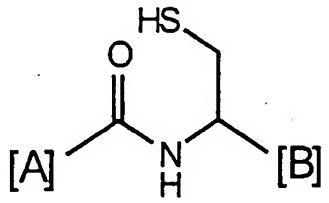
where TRFD is a thiol reactive function derivative and the Spacer is a linking group,

comprising the steps of:

(i) forming an intermediate compound having the formula (II):

[A] — L' — [B] (II)

wherein intermediate linker group L' comprises a free thiol group, and wherein the intermediate compound of formula (II) has the chemical structure shown in formula (VII):



(VII)

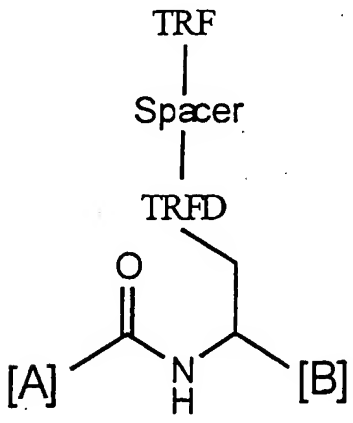
by reacting the functional group shown in formula (III) of native [A] to the terminal functional group shown in formula (IV) of native [B] using native chemical ligation;

(ii) reacting a spacer linker having the formula (VIII):

TRF-Spacer-TRF (VIII)

where TRF is a thiol reactive function (TRF) group and the Spacer is a linking group,

with the free thiol group of intermediate linker group L' to form a second intermediate compound having the formula (XII):

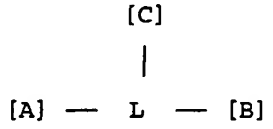


(XII)

and

(iii) producing the compound having the formula (I) by reacting the thiol of native [C] to the thiol reactive function (TRF) group of the second intermediate compound having the formula (XII).

Alternatively, there is provided a process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

(iv) admixing the second intermediate compound with a thiol reactive function (TRF) group linker having a thiol reactive function (TRF) group and a free thiol in a reaction solution;

(v) allowing the thiol reactive function (TRF) group of the thiol reactive function (TRF) group linker to react with the free thiol group of the second intermediate compound to form a third intermediate compound with a free thiol;

(vi) admixing native [C] with the third intermediate compound in a reaction solution; and

(vii) reacting the thiol reactive function (TRF) group of native [C] with the free thiol of the third intermediate compound for producing a compound having the formula (I).

The thiol reactive function (TRF) group linker may comprise, or the Spacer may be, a polyalkyloxy, alkyl, aryl, arylalkyl or peptidyl group.

Each of R_1 and R_2 shown in formula (III) and formula (IV) may be a substituted or unsubstituted aryl group or a substituted or unsubstituted alkyl group.

The functional group of native [B] may be a 1-amino-2-thiol group having an unoxidised sulfhydryl side chain.

The unoxidised sulfhydryl side chain of the 1-amino-2-thiol group of native [B] may be the only unoxidised sulfhydryl side chain

present in native [B]. Alternatively, native [B] may comprise additional free thiols which are removed by mutation such that the unoxidised sulfhydryl side chain of the 1-amino-2-thiol group of native [B] becomes the only unoxidised sulfhydryl side chain present.

Native [B] may have a terminal cysteine moiety comprising the terminal 1-amino-2-thiol group.

The thiol reactive function (TRF) group may be a thiol, an alkyl halide, a 2-pyridyl disulfide, a 4-pyridyl disulfide or a Michael acceptor (eg. a vinyl sulfone or a maleimide).
An example of a suitable thiol reactive function (TRF) is an alkyl halide and the reaction is an S_N2 displacement. The reaction product is a thioether. Incorporation of alkyl halides into various substrates is a straightforward process well known to those skilled in the art.

Another example is an activated disulfide such as a 2- or 4-pyridyl disulfide and the reaction is a disulfide exchange. The reaction product is a disulfide. Pyridyl disulfides may be incorporated synthetically into many substrates. Additionally, a range of reagents is available to install pyridyl disulfides on the surface of proteins and other molecules.

A further example is a Michael acceptor such as a vinyl sulfone or a maleimide and the reaction is a Michael-type addition. The reaction product is a thioether. Maleimides may be incorporated synthetically into many substrates. Additionally, a range of reagents is available to install maleimides on the surface of proteins and other molecules.

Examples of the three resultant reaction products are shown (Figure 3). Other chemical functions are able to react chemically with free thiols and are well known to those skilled in the art (for example, BMOE).

The thioester group or functional group of native [A] may be a benzyl, ethyl or 2-aminoethyl group.

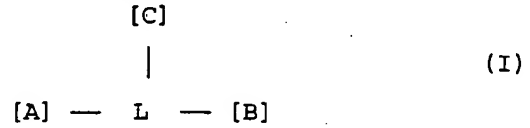
Native [A] and/or native [B] and/or native [C] may comprise any one of the following entities or derivatives thereof: pharmacophore, ligand, small molecule, purification handle/immunochemical tag (eg biotin, a dinitrophenyl compound), fluorescent moiety (eg. Alexafluor dye, Texas Red dye), solubilising agent, chelating ligand, chelating ligand plus radioimaging agent, therapeutic protein, antibody or fragment thereof, peptide, peptidomimetic or macroscopic particle (eg. a bead).

Prior art (for example US5789653) does not suggest or teach a covalently linked combination therapeutic covalently linked to a directing ligand or antibody or antibody fragment.

Native [A] and/or native [B] and/or native [C] may comprise an entity from a combinatorial array of chemicals. The combinatorial array may be a peptide library.

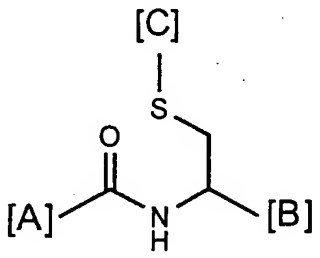
Also provided according to the present invention is compound obtainable or obtained by the above-mentioned processes.

Further provided is a compound having the general formula (I):



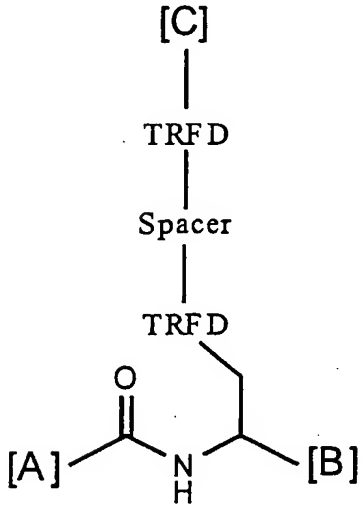
(I)

wherein [A], [B] and [C] are derived from chemical entities native [A], native [B] and native [C], respectively, and are covalently linked by linker group L, the compound of formula (I) having a chemical structure shown in formula (VI):



(VI)

or formula (XI):



(XI)

where TRFD is a thiol reactive function derivative and the Spacer is a linking group,

and wherein each of [A], [B] and [C] comprises any one of the following entities or derivatives thereof: pharmacophore,

ligand, small molecule, purification handle/immunochemical tag (eg biotin, a dinitrophenyl compound), fluorescent moiety (eg. Alexafluor dye, Texas Red dye), solubilising agent, chelating ligand, chelating ligand plus radioimaging agent, therapeutic protein, antibody or fragment thereof, peptide, peptidomimetic or macroscopic particle (eg. a bead).

Yet further provided is the use of the compound according to the present invention in a binding screen for detecting binding to a target entity. The binding screen may select compounds which have increased or decreased binding to a target entity.

Also provided is the use of a compound formed by native chemical ligation in a process for the production of the compound of the invention.

In one embodiment, native [A] and/or native [B] and/or native [C] may comprise a myristoyl/electrostatic switch peptide (MSWP) or derivative thereof, for example MSWP2278 (SEQ. ID NO: 2).

In another embodiment, native [A] comprises EGFP, native [B] comprises MSWP2278 (SEQ. ID NO: 2) and native [C] comprises Texas Red Dye.

In yet another embodiment, native [A] comprises EGFP, native [B] comprises MSWP2278 (SEQ. ID NO: 2), and native [C] comprises an antibiotic (for example vancomycin).

In an alternative embodiment, native [A] comprises Hirulog anticoagulant peptide, native [B] comprises MSWP2278 (SEQ. ID NO: 2), and native [C] comprises SCR1-3.

WO 98/02454 relates to the derivatisation of soluble complement regulators by disulfide exchange reaction with pyridyl disulfide activated myristoyl/electrostatic switch peptides (MSWPs). These derivatives demonstrate improved complement inhibitory activity owing to their membrane localisation by the MSWP. The myristoyl group was designed to be minimally membrane insertive, while the lysine rich peptidic portion of MSWP-1 (Seq. ID No 1) was designed to interact with the negatively charged phospholipid headgroups present on the lipid bilayer. The concept is exemplified in WO 98/02454 by SCR1-3 of LHR-A of CR1, altered by a single C-terminal cysteine residue. After treatment with an appropriate reducing agent to ensure the presence of a free thiol on the terminal cysteine sidechain, the protein is reacted with MSWP-1 (N-(Myristoyl)-Gly-Ser-Ser-Lys-Ser-Pro-Ser-Lys-Lys-Lys-Lys-Lys-Lys-Pro-Gly-Asp-(S-2-Thiopyridyl)Cys-NH_2) to yield a membrane binding derivative of SCR1-3 (Figure 4). The resultant conjugate protein showed anti-haemolytic activity orders of magnitude greater than that displayed by the underivatised protein alone.

This chemistry relies on there existing within the protein participant of the reaction a single free thiol resulting in derivatisation at a single point only. There are numerous other cysteines in SCR1-3, but all save the C-terminal cysteine are participants in intramolecular disulfide bridges. Some proteins, however, may have free thiols other than that specific free thiol on which derivatisation is desired to occur. In instances such as this, it may be possible to mutate using molecular biological techniques these cysteine residues to other amino acids, while retaining the desired biological function. In other cases this mutation process may result in the loss of the desired biological function.

WO 98/02454 describes polypeptide derivatives in terms of a combinatorial array of membrane binding elements corresponding to the general structure:

$$[P]-\{L-[W]\}_n-X \qquad 1$$

in which:

P is the soluble polypeptide,

each L is independently a flexible linker group,

each W is independently a peptidic membrane binding element,

n is an integer of 1 or more, and

X is a peptidic or non-peptidic membrane-binding entity which may be covalently linked to any W.

In the context of the current invention where such an assembly is constructed by native chemical ligation, and where n = 1, structure 1 can be rewritten whereby entity A represents [P], entity B represents [W] and X, and thus L is comprised of the nascent cysteine residue at the junction between A and B. We will for simplicity continue our discussion of the scope of the invention using A, B, and C, but with an additional linker function L which is comprised of the nascent cysteine residue at the junction between A and B and any further atoms that may make up a flexible linker to moiety C.

$$\begin{array}{c}[C]\\ |\\ [A]-L-[B]\end{array} \qquad 2$$

Where [A] is a protein, one method of formation of the [A]-L-[B] portion of the invention is provided by the IMPACT™ system (New England Biolabs). Conjugates are afforded by the native chemical ligation of peptides containing an N-terminal cysteine, to proteins expressed in frame with a controllable intervening peptide sequence (CIVPS) and an affinity purification handle. The desired protein is cleaved from its affinity purification matrix by elution with a free thiol such as MESNA to yield a C-terminal thioester. The MESNA thioester is displaced by the thiol of the N-terminal cysteine before an S→N acyl shift installs the peptide bond in an irreversible fashion. Alternatively, the affinity matrix is incubated with the free peptide, displacing the desired protein from the matrix and ligating in a single step (Figure 5).

The most favoured MSWP in WO 98/02454 comprised an N-terminal myristoyl function, a C-terminal 2-thiopyridylated cysteine primary amide, and a series of intervening residues including the hexalysine region (Figure 4). The ligation described above required an N-terminal cysteine not present in MSWP-1. MSWP-1 was modified for this purpose to include an N-terminal lysine primary amide modified at its ε-amino function to incorporate a surrogate N-terminal cysteine. This participates in the native chemical ligation while retaining the overall directionality of the so formed tail. This peptide is described as MSWP2278 (Seq. ID No 2). The principle demonstrated by MSWP2278 whereby a peptide may participate in a native chemical ligation *via* the intermediacy of a 1-amino-2-thiol engineered onto any sidechain using standard solid phase techniques is contained within the scope of the invention.

In the specific case where [C] = [A]-L-[B] it is possible to form a symmetrical homodimer. The alkyl halide, activated

disulfide, and maleimide methods described above are also applicable in this case to yield symmetrical thioethers, symmetrical disulfides, and symmetrical maleimides respectively. However a further method also applies. The invention allows for the spontaneous air oxidation of two identical [A]-L-[B] functions to the homodimer in aqueous solution.

3

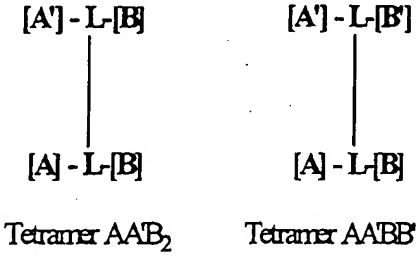
Many other linking chemistry are possible. Some are explained in WO 98/02454 and are incorporated herein by reference.

Clearly if one or more of the following:

$$A \neq A'$$

$$B \neq B'$$

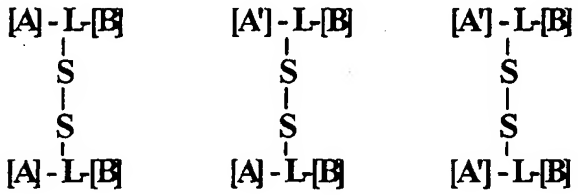
is true, then the product of such a coupling reaction will be pseudoheterotetra- or heterotetrameric i.e.



Tetramer $AA'B_2$ Tetramer AA'BB'

4

In both of these cases, a simple air oxidation may result in mixtures of products formed, therefore the activated disulfide and Michael type addition methods are preferred. For instance, in the case of the air oxidation of [A]-L-[B] + [A']-L-[B], the following three products would be observed in the reaction mixture:



The incorporation of C orthogonally to the existing linear molecular backbone is a fundamental utility described within the invention. The entity C is, as a result of the inventive step, not constrained within the confines of a linear sequence but is attached to a point between the entities A and B.

In the case where one of A, B, or C is a probe such as biotin, or a fluorophore such as an Alexafluor dye, an immunochemical marker such as a dinitrophenyl compound, or a radiochemical label, the trimers or tetramers described above gain utility in that their fate is able to be monitored. These monitoring techniques include but are not limited to an immunohistochemical screen, a fluorescence localisation assay, or radiochemical screen. The scale of monitoring may vary from a whole body screen, to intra-organ differential localisation, to the identification of sub-cellular fate.

An example is:

A = EGFP

B = MSWP2278

C = Texas Red Dye

This heterotrimer (Figure 6) contains a green fluorescent protein directed towards cell membranes by the myristoyl electrostatic switch peptide. A Texas Red dye was incorporated as [C] using the appropriate maleimide reagent (Molecular Probes - Eugene, OR). This heterotrimer under normal neutral pH conditions in aqueous media fluoresces in both the red and green channels. When incubated with cells, the trimer localises to the cell membrane, where both types of fluorescence may be observed. Some of the heterotrimer passes through the cell membrane and localises to the lysosomal compartments. The low pH found in these compartments abrogates the green fluorescence associated with EGFP, leaving only the Texas Red to fluoresce. This molecule, therefore, can be used as a lysosomal tracking device. Many other applications such as this will become apparent to those skilled in the art through the teachings of this invention.

A further example is:

A = EGFP

B = MSWP2278

C = An antibiotic

The antibiotic in can be vancomycin for example (Figure 7).

A further example is:

A = Hirulog anticoagulant peptide

B = MSWP2278

C = SCR1-3

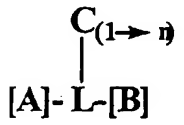
This heterotrimer contains all the elements of a ligand directed combination therapy as described earlier. The MSWP2278 peptide serves to localise the remainder of the molecule on the surface of mammalian cells. On these surfaces, the antiocoagulant properties and the complement inhibitory activities of both of the components **A** and **C** respectively are conserved. The synthesis was accomplished by the native chemical ligation of MSWP2278 with a hirulog type peptide thioester ((D)-PheProArgProGlyGlyGlyGlyAspGlyAspPheGluGluIleProGluGluTyrLeuGlySerGlySerGlySerGlySerGlySerGlySerGly-SEtCONH$_2$) (Seq ID No 3). The product formed was reacted with a bismaleimide, and then with SCR1-3. The single exposed C-terminal cysteine reacts solely with the maleimide to provide the heterotrimer (Figure 8). The complement inhibition activity observed for SCR1-3 linked to MSWP1 was almost completely retained, in addition to an anticoagulant activity associated with a membrane directed hirulog type peptide. The combination of a membrane-directed complement inhibitor and an anti-coagulant in a single therapeutic agent may be useful for applications such as transplantation.

In a further example, **C** can be incorporated as described above not as a single chemical entity but as a combinatorial array of n chemical entities where **C** = **C(1→n)**.

Alternatively, this combinatorial array of n chemicals was comprised of a peptide library.

Further, each of n chemical entities was comprised of a different peptide sequence, and was linked to **L** by disulfide exchange chemistry. This principle permitted the use of the combinatorial

array of C(1→n) in a binding screen where the binding energy gained from C was additive and augments that gained from B. The process of identifying binding ligands identifies C sequences capable of increasing the overall binding to the cell type of choice, and thus confers selectivity over other cell types



6

i.e. for library member λ:

$\Delta G\ (\lambda) = \Sigma\ (\Delta G_A + \Delta G_B + \Delta G_C)$

And in the case where B is a MSWP, $\Delta G_B = \Delta G_W + \Delta G_X$.

There are several possibilities to enable the reduction to practice of such a binding screen. One method involved a one-pot preparation of a multiplicity of library members which were incubated with a tissue for which selective targeting was desired. Each member of this library shall have incorporated into C an affinity handle such as biotin, and shall have a disulfide linkage between C and L. After a washing process to remove library members that either do not bind or bind weakly, the sequence identity of C ligands that bind can be determined by reduction of the disulfide, extraction of C by supported avidin, and identification of C by mass spectrometry.

A different method involves the combinatorial assembly of heterotrimers where the members of the library C(1→n) are used

individually in the fashion described above to furnish a library of discrete heterotrimers in a high throughput fashion. The members of this protein library can then be individually screened for binding to a particular cell type in, for instance, a 96-well format. The extent of binding can be detected immunohistochemically (ELISA) or by direct fluorescence in the case where **A** is fluorescent such as GFP, or where **A** or **B** or **C** or any combination of the three is a fluorescently labeled entity. Both of these methods are exemplified (*vide infra*).

Any of the components brought together with this enabling technology may be selected from the group consisting of: a pharmacophore, a ligand, a small molecule, a purification handle/immunochemical tag (eg biotin), a fluorescent moiety, a solubilising agent, a chelating ligand, a chelating ligand plus radioimaging agent, a therapeutic protein, an antibody or fragment thereof, a peptide or a peptidomemtic.

This list is not exhaustive. Other moieties may be linked in the fashion to be described and will be apparent to those skilled in the art.

Drawings

Figure 1: Native Chemical Ligation schematic.

Figure 2: Formation of a heterotrimer by reaction of a NCL product with a thiol-reactive component.

Figure 3: Examples of heterotrimeric products.

Figure 4: Preparation of a membrane binding derivative of SCR1-3.

Figure 5: Ligation of a MSWP to an intein expressed protein.

Figure 6: Two channel fluorescence, pH dependent heterotrimer.

Figure 7: Membrane-targeted antibiotic/fluorescent protein heterotrimer.

Figure 8: Membrane-targeted Hirulog anticoagulant/SCR1-3 heterotrimer.

Figure 9: Binding of untailed and tailed EGFP to COS7 cells. A: Fluorescence micrograph after incubation with EGFP. B: Fluorescence micrograph after incubation with APT2308.

Examples

Example 1: Synthesis of disulfide exchange MSWP1 (SEQ ID NO: 1)

The peptide:
Myristoyl-Gly-Ser-Ser-Lys-Ser-Pro-Ser-Lys-Lys- Lys-Lys-Lys-Lys-Pro-Gly-Asp-Cys-(S-2-thiopyridyl)-NH_2 was synthesised according to the method described in WO 98/02454.

Example 2: Synthesis of Native Chemical Ligation/MSWP2278 (SEQ ID NO: 2)

The peptide:
Myristoyl-Gly-Ser-Ser-Lys-Ser-Pro-Ser-Lys-Lys-Asp-Asp-Lys-Lys-Pro-Gly-Asp-(ε-amino-cysteinyl carboxamide)-Lys-NH_2 was prepared by solid phase synthesis using Boc synthesis on MBHA resin (Nova). Coupling reactions were carried out using appropriately protected Boc amino acid monomers (Nova) activated with TBTU aided with HOBT (Alexis Bio. Co.) with ninhydrin monitoring after

each extension. The first amino acid was installed as α-amino-Boc, ε-amino-Fmoc lysine, and the Fmoc protection then removed with 20% piperidine. α-Amino-Fmoc, S-methoxybenzyl cysteine was then coupled to the ε-amino group before the remainder of the synthesis was carried out using appropriately protected Boc monomers. Cleavage from the resin and cleavage of the sidechain protecting groups was accomplished with high HF conditions using p-cresol and p-thiocresol as scavengers at 0°C over 1 h.

The crude peptide was desalted *via* gel filtration (Sephadex G10, 0.1% TFA in water) before purification by preparative C18 Vydac) high performance liquid chromatography (HPLC) using 0.1% trifluoroacetic acid and 0.1% trifluoroacetic acid/90% acetonitrile as gradient components. After lyophilisation, the peptide was a white amorphous powder, soluble to at least 10 mg/mL in both dimethylsulfoxide and water. Fast atom bombardment spectrometry gave a main peak at 2101 amu corresponding to the molecular ion of the peptide.

Example 3: Expression of recombinant EGFP in the intein system

Standard molecular biological techniques were used to replace the VP1 gene in vector pET21-VP1Int [Schmidt U., Rudolph, R. & Böhm, G. *J. Virol.* 74, 1658-1662 (2000)] with the gene for enhanced green fluorescent protein (EGFP), using the vector pEGFP-N1 (Clontech) as template. The resulting vector, pET107-01, expresses the EGFP gene under the control of the T7*lac* promoter as a C-terminal fusion with the VMA1 gene for the intein from *Saccharomyces cerevisiae* and the gene for the chitin binding domain from *Bacillus circulans*. The construction of pET21-VP1Int from vectors pCYB2 and pET21d (New England Biolabs) has been described elsewhere (Schmidt *et al.*, *vide supra*).

Upon transformation into the bacterial strain Hams113, an overnight starter culture of 25 mL LB medium containing 100 µg/mL ampicillin was inoculated with a single colony and, on the next morning, diluted 1:100 into 2 L of LB medium containing 100 µg/mL ampicillin. The culture was fermented at 37°C until an optical density at 600 nm of 0.5 was achieved, at which point the temperature was lowered to 20°C and expression of the gene induced by the addition of 1 mM IPTG after 30 mins. Cells were harvested 21 h post induction by low-speed centrifugation and stored in aliquots at -40°C.

Example 4: Affinity purification of EGFP and tailing with APT542

To lyse cells, an aliquot of bacterial paste corresponding to 333 mL of fermentation broth was resuspended in 40 mL of ice-cold Buffer 100 (20 mM HEPES pH 8.0, 100 mM NaCl, 1 mM EDTA) and homogenised by three passages through an Emulsiflex C5 homogeniser at a pressure of 12 kpsi. The soluble fraction was obtained by centrifugation (20000 g, 30 min, 4°C) and immediately passed over a column containing 15 mL chitin beads (New England Biolabs). The column was washed with four column volumes of Buffer 100, eight column volumes of Buffer 2000 (20 mM HEPES pH 8.0, 2000 mM NaCl, 1 mM EDTA), four column volumes of Buffer 100, and finally with three column volumes of Cleavage Buffer 1 (Buffer 100 containing 50 mM cysteine). After incubation at 6°C for 20 h, EGFP was eluted from the column using Buffer 100 as the elution buffer.

The procedure outlined here resulted in an EGFP labelled at the C-terminal with a single additional cysteine residue designated

APT2269. The availability of this thiol was tested by derivatisation with the thiol-reactive MSWP1. 500 µL of a 26 µM solution of APT2269 was mixed with 100 µL 10 mM TCEP/20 mM HEPES over 1 h to ensure complete reduction of the terminal cysteine. The mixture was dialysed against 1 L PBS/EDTA (1 mM)/TCEP (1 mM), against 1 L PBS/EDTA (1 mM) and MSWP1 added to a final concentration of 34 µM. This reaction resulted in the derivatisation of APT2269 with between 1 and 3 tails (a process referred to as supertailing).

Example 5: Expression of recombinant EGFP in the intein system and tailing with MSWP2278 to yield APT2308

EGFP was purified as described in Example 4, but Cleavage Buffer 2 (Buffer 100 containing 50 mM MESNA and 50 µM APT2278) was used in the overnight incubation. The cleavage resulted in almost quantitative tailing of EGFP with no evidence of supertailing. Even upon addition of 5 mM APT2278 (equivalent to a 1000-fold molar excess) no further tailing was detected. The reaction product of EGFP with APT2278 was designated APT2308.

Example 6: Further derivatisation of APT2308 with APT542

To demonstrate the accessibility of the newly introduced thiol in APT2308, we subjected it to further derivatisation with the thiol-reactive compound APT542.

APT2308 was treated with 1 mM TCEP overnight and the buffer exchanged by passing the solution over a Nap-5 column equilibrated with PBS. APT542 was added to a concentration of 13 µM and the reaction mixture analysed by SDS-PAGE after 1 h.

The addition of each tail was confirmed by a 2 kDa shift in molecular weight.

Example 7: Binding of APT2308 to COS7 cells

COS7 cells were incubated with either APT2308 or recombinant EGFP at a concentration of 6 μM in Dulbecco's modified eagle medium (Sigma). After 1 h, the cells were visualised by fluorescence microscopy and the images recorded with identical parameters. EGFP demonstrated no binding to COS7 cells (which had been demonstrated with other cell lines for concentrations up to 2.5 mg/mL). APT2308 bound to the cells with very high efficiency (Figure 9).

Example 8: Mutation of EGFP to remove two unpaired cysteines and expression in the intein system - APT2307

pET107-01 was used as a template for mutagenesis with the aim of replacing two cysteine residues in APT 2253 (C49 and C71), with two serine residues. Single base changes (TGC to AGC) were made to each codon at positions 5349 and 5415 of pET107-01.

Mutagenesis primers were designed to extend 15 bases in either direction of each mismatched base. A restriction site was also incorporated into each pair of primers to aid the identification of positive clones. In both cases, this necessitated the introduction of a silent mutation in the coding sequence.

Primers Designed for the Mutation of Cysteine Residue at Position 49 (C49S)

Primer 1200951: 5' CC CTG AAG TTC ATC AGC ACC ACC GGT AAG CTG CCC GTG CCC 3'

Primer 1200952: 5' GGG CAC GGG CAG CTT ACC GGT GGT GCT GAT GAA CTT CAG GG 3'

In addition to generating a thymine to adenine transition (at position 5349 in pET107-01), these 41mers incorporate an Age I site (ACCGGT) by silent mutation of a glycine codon (G52 in APT 2253), replacing cytosine with thymine (position 5360 in pET107-01).

Primers Designed for the Mutation of Cysteine Residue at Position 71 (C71S).

Primer 1200961: 5' G ACC TAC GGC GTG CAA AGC TTC AGC CGC TAC C 3'

Primer 1200962: 5' G GTA GCG GCT GAA GCT TTG CAC GCC GTA GGT C 3'

In addition to generating a thymine to adenine transition (at position 5415 in pET107-01), these 32 mers incorporate a Hind III site (AAGCTT) by silent mutation of a glutamine codon (Q70 in APT 2253), replacing guanine with adenine.

A two step process was used whereby the C49S variant was generated first via site-directed mutagenesis using the QuickChange mutagenising kit (Stratagene), according to the manufacturers instructions. pET107-01 was amplified using primers 1200951 and 1200952, followed by treatment with Dpn I and transformation into XL10 competent cells (Stratagene) to generate pET107-02. pET107-02 was then used as a template for C71S

mutagenesis with primers 1200961 and 1200962, using the same process. The resultant plasmid, pET107-04, encodes APT 2307.

MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFIST TGKLPVPWPT LVTTLTYGVQ SFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYKP GCFAKGTNVL MADGSIECIE NIEVGNKVMG KDGRPREVIK LPRGRETMYS VVQKSQHRAH KSDSSREVPE LLKFTCNATH ELVVRTPRSV RRLSRTIKGV EYFEVITFEM GQKKAPDGRI VELVKEVSKS YPISEGPERA NELVESYRKA SNKAYFEWTI EARDLSLLGS HVRKATYQTY APILYENDHF FDYMQKSKFH LTIEGPKVLA YLLGLWIGDG LSDRATFSVD SRDTSLMERV TEYAEKLNLC AEYKDRKEPQ VAKTVNLYSK VVRGNGIRNN LNTENPLWDA IVGLGFLKDG VKNIPSFLST DNIGTRETFL AGLIDSDGYV TDEHGIKATI KTIHTSVRDG LVSLARSLGL VVSVNAEPAK VDMNGTKHKI SYAIYMSGGD VLLNVLSKCA GSKKFRPAPA AAFARECRGF YFELQELKED DYYGITLSDD SDHQFLLGSQ VVVHACGGLT GLNSGLTTNP GVSAWQVNTA YTAGQLVTYN GKTYKCLQPH TSLAGWEPSN VPALWQLQ.

Example 9: Tailing of APT2307 with APT2278 to yield APT2360, with APT2335 to yield APT2361, and with cysteine to yield APT2359

Doubly mutated EGFP was tailed using the appropriate peptide or amino acid at 50 μM as described in Example 5. Each reaction mixture was ultrafiltered exhaustively with PBS to remove excess peptide or amino acid.

Example 10: Derivatisation of APT2360 with Biotin - APT2396

APT2360 (40 μl of a 10 μM solution in PBS) was mixed with TCEP (2.5 μl of a 1 mM solution in 50 mM HEPES pH 4.5) and left at ambient temperature overnight. EZ-Link PEO-maleimide biotinylation reagent (Pierce, UK). 8 μL of a 400 uM solution in PBS) was added and the mixture left at ambient temperature

over 1 h. Freshly made cysteine solution (5 μL of a 10 mM solution in PBS) was added to quench the reaction. Excess reagents were removed by microdialysis. A small gel shift was observed, the new product was immunoreactive with an anti-biotin antibody in a Western Blot, and the expected increase in molecular mass was observed by MALDI mass spec. The reaction proceeded in quantitative yield.

Example 12: Derivatisation of APT2360 with an antibiotic - APT2398

APT2360 was treated with TCEP as described in Example 11 before an activated disulfide derivative of vancomycin (APT2033), at 13.2 μL of a 610 μM solution in PBS, was added. After 2 h at ambient temperature, excess reagents were removed by microdialysis. A significant gel shift was observed, and the new product was formed quantitatively.

Example 13: Derivatisation of APT2361 with a biotinylated peptide- APT2404

APT2361 was treated with TCEP as described for APT2360 in Example 11 before a 3-fold molar equivalency of $^{-}$OOC-MTAAPLRGSXGC(2-S-Py)-NH2 (X=biotinylated lysine) (APT2400) was added. The mixture was incubated at ambient temperature over 2 h, before excess TCEP and APT2400 were removed by dialysis. An appropriate gel shift was observed by SDS-PAGE corresponding to the addition of a single peptide to the native chemical ligation junction.

Example 14: Derivatisation of APT2360 with Texas Red C2 maleimide - APT2455

APT2360 was prepared as described in Example 9. Upon incubation overnight in the presence of 1 mM TCEP, the sample was dialysed against degassed PBS and Texas Red-maleimide (Molecular Probes) was added to a final concentration of 2 mM (corresponding to an approximate 200-fold excess over APT2360). The mixture was incubated for 2 h at room temperature and then the protein separated from free fluorophore via gel filtration with a PD-10 column. An analysis using SDS-PAGE, UV/VIS spectroscopy and cell binding studies showed that APT2455, the reaction product

- absorbed with both fluorophores in neutral to alkaline pH
- absorbed only with the Texas Red Fluorophore in acidic pH
- bound to cells and caused green and red surface staining as well as predominantly red vesicular staining, due to uptake into acidic lysosomes.

Example 15: A membrane-targeted anticoagulant and complement inhibitor - APT2434

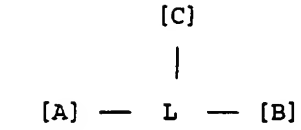
The production of APT2434 takes place in a three-step synthesis. The first step is the Native Chemical Ligation of APT2383, the hirulog-thioester (including a Glycine-Serine linker segment) to APT2278, yielding APT2384. Typically, both reaction partners are used at mM concentrations and the reaction proceeds overnight at 22 degrees centigrade in the presence of 1 mM TCEP and in 40 mM HEPES buffer, pH 8.5. APT2384 was purified using RP-HPLC and derivatised with a tenfold surplus of BM(PEO)4 in PBS, a bismaleimide with an 18 Å long linker segment (Pierce). The reaction product, APT2433, was purified via RP-HPLC and reacted with APT154 to yield APT2434, the final heterotrimer.

APT2434 was assessed regarding its three proposed functions - membrane localisation, complement inhibition and anticoagulant

activity. All three functions were found to be preserved in the final molecule.

Claims

1. A process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

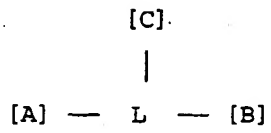
(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

(iv) admixing native [C] with the second intermediate compound in a reaction solution; and

(v) reacting the thiol reactive function (TRF) group of native [C] with the free thiol group of the second intermediate compound for producing a compound having the formula (I).

2. A process for the synthesis of a compound having the formula (I):



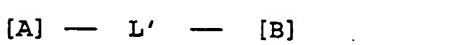
(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) reacting the thioester group of native [A] to the 1-amino-2-thiol group of native [B] by native chemical ligation to form an intermediate compound having the formula (II):



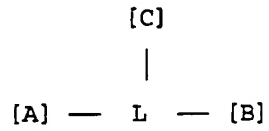
(II)

where intermediate linker group L' comprises an amide bond having attached thereto a free thiol;

and

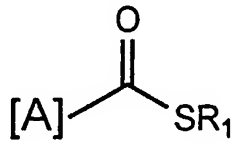
(ii) reacting the thiol reactive function (TRF) group of native [C] to the free thiol of the intermediate compound having the formula (II) to form the compound having the formula (I).

3. A process for the synthesis of a compound having the formula (I):



(I)

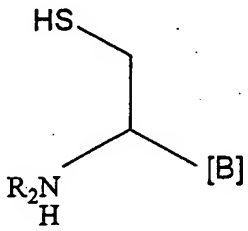
from chemical entities native [A], native [B] and native [C], where native [A] comprises a functional group having a chemical structure, excluding [A], shown in the formula (III):



(III)

where R_1 is H or a sidechain,

native [B] comprises a functional group having a chemical structure, excluding [B], shown in the formula (IV):



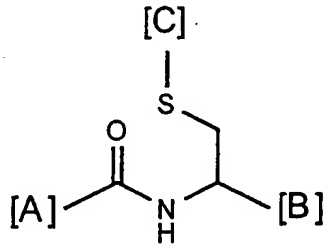
(IV)

where R_2 is H or a sidechain,

and native [C] comprises a thiol reactive function (TRF) group attached to [C] as shown in formula (V):

(V)

wherein [A], [B] and [C] are chemical entities covalently linked by a linker group L, and wherein the compound of formula (I) has the chemical structure shown in formula (VI):



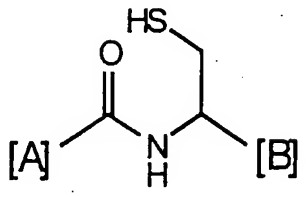
(VI)

comprising the steps of:

(i) forming an intermediate compound having the formula (II):

[A] — L' — [B] (II)

wherein intermediate linker group L' comprises a free thiol group, and wherein the intermediate compound of formula (II) has the chemical structure shown in formula (VII):



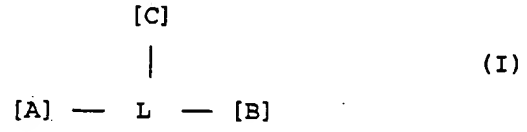
(VII)

by reacting the functional group shown in formula (III) of native [A] to the functional group shown in formula (IV) of native [B] using native chemical ligation;

and

(ii) producing the compound having the formula (I) by reacting the thiol reactive function (TRF) group of native [C] to the free thiol of the intermediate linker group L'.

4. A process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

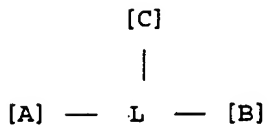
(iv) admixing the second intermediate compound with a thiol reactive function (TRF) group linker having at least two thiol reactive function (TRF) groups in a reaction solution;

(v) allowing the thiol reactive function (TRF) group linker to react with the free thiol group of the second intermediate compound to form a third intermediate compound with a thiol reactive function (TRF) group;

(vi) admixing native [C] with the third intermediate compound in a reaction solution; and

(vii) reacting the thiol group of native [C] with the thiol reactive function (TRF) group of the third intermediate compound for producing a compound having the formula (I).

5. A process for the synthesis of a compound having the formula (I):



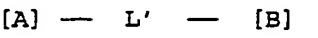
(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

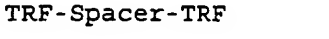
(i) reacting the thioester group of native [A] to the 1-amino-2-thiol group of native [B] by native chemical ligation to form a first intermediate compound having the formula (II):



(II)

where intermediate linker group L' comprises an amide bond having attached thereto a free thiol group;

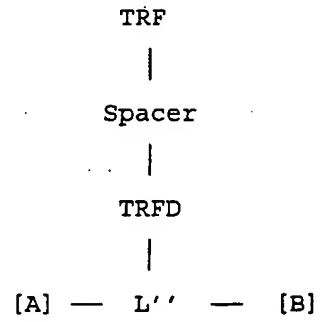
(ii) reacting a spacer linker having the formula (VIII):



(VIII)

wherein TRF is a thiol reactive function (TRF) group and the Spacer is a linking group,

with the free thiol of intermediate linker group L' to form a second intermediate compound having the formula (IX):



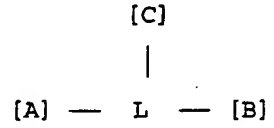
(IX)

where TRFD is a thiol reactive function derivative attached to second intermediate linker group L'',

and

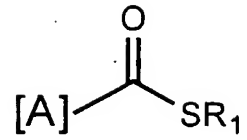
(iii) reacting the thiol of native [C] to the thiol reactive function (TRF) group of the second intermediate compound having the formula (IX) to form the compound having the formula (I).

6. A process for the synthesis of a compound having the formula (I):



(I)

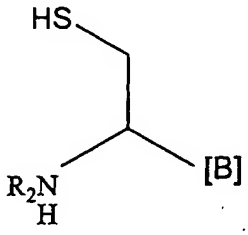
from chemical entities native [A], native [B] and native [C], where native [A] comprises a functional group having a chemical structure, excluding [A], shown in the formula (III):



(III)

where R_1 is H or a sidechain,

native [B] comprises a functional group having a chemical structure, excluding [B], shown in the formula (IV):



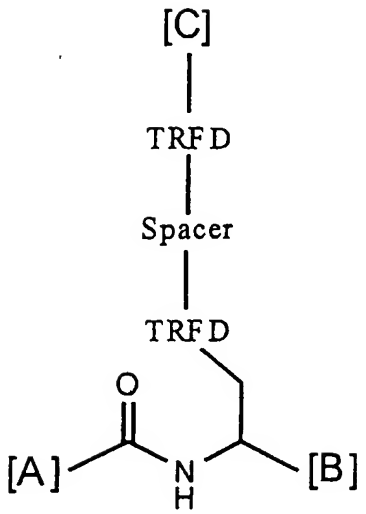
(IV)

where R_2 is H or a sidechain,

and native [C] comprises a thiol attached to [C] as shown in formula (X):

(X)

wherein [A], [B] and [C] are chemical entities covalently linked by a linker group L, and wherein the compound of formula (I) has the chemical structure shown in formula (XI):



(XI)

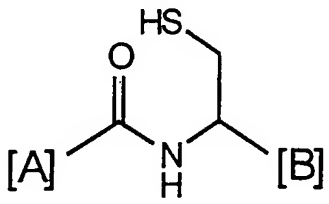
where TRFD is a thiol reactive function derivative and the Spacer is a linking group,

comprising the steps of:

(i) forming an intermediate compound having the formula (II):

[A] — L' — [B] (II)

wherein intermediate linker group L' comprises a free thiol group, and wherein the intermediate compound of formula (II) has the chemical structure shown in formula (VII):



(VII)

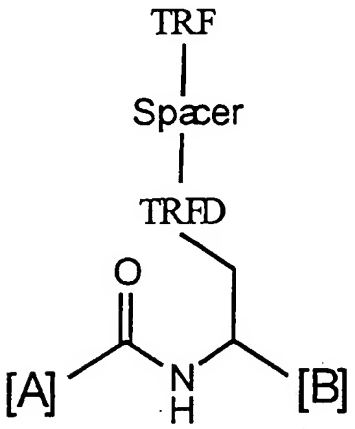
by reacting the functional group shown in formula (III) of native [A] to the terminal functional group shown in formula (IV) of native [B] using native chemical ligation;

(ii) reacting a spacer linker having the formula (VIII):

TRF-Spacer-TRF (VIII)

where TRF is a thiol reactive function (TRF) group and the Spacer is a linking group,

with the free thiol group of intermediate linker group L' to form a second intermediate compound having the formula (XII):

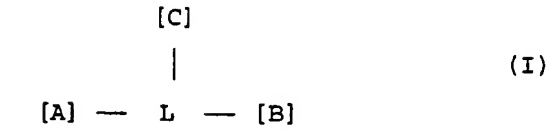


(XII)

and

(iii) producing the compound having the formula (I) by reacting the thiol of native [C] to the thiol reactive function (TRF) group of the second intermediate compound having the formula (XII).

7. A process for forming a compound having the formula (I):



(I)

from chemical entities native [A], native [B] and native [C], where native [A] has a thioester group, native [B] has a 1-amino-2-thiol group with an unoxidised sulfhydryl side chain, and native [C] has a thiol reactive function (TRF) group,

wherein chemical entities [A], [B] and [C] are covalently linked by linker group L,

comprising the steps of:

(i) admixing native [A] and native [B] in a reaction solution;

(ii) condensing the unoxidised sulfydryl side chain of native [B] with the thioester group of native [A] for producing a first intermediate compound wherein [A] and [B] are linked with a β-aminothioester bond;

(iii) rearranging the β-aminothioester bond for producing a second intermediate compound wherein [A] and [B] are linked with an amide bond having attached thereto a free thiol group;

(iv) admixing the second intermediate compound with a thiol reactive function (TRF) group linker having a thiol reactive function (TRF) group and a free thiol in a reaction solution;

(v) allowing the thiol reactive function (TRF) group of the thiol reactive function (TRF) group linker to react with the free thiol

group of the second intermediate compound to form a third intermediate compound with a free thiol;

(vi) admixing native [C] with the third intermediate compound in a reaction solution; and

(vii) reacting the thiol reactive function (TRF) group of native [C] with the free thiol of the third intermediate compound for producing a compound having the formula (I).

8. The process according to any one of claims 4-7, wherein the thiol reactive function (TRF) group linker comprises, or the Spacer is, a polyalkyloxy, alkyl, aryl, arylalkyl or peptidyl group.

9. The process according to either one of claim 3 or claim 6, wherein each of R_1 and R_2 is a substituted or unsubstituted aryl group or a substituted or unsubstituted alkyl group.

10. The process according to any one of claims 3, 6 or 9 wherein the functional group of native [B] is a 1-amino-2-thiol group having an unoxidised sulfhydryl side chain.

11. The process according to any one of claims 1, 2, 4, 5, 7 or 10 wherein the unoxidised sulfhydryl side chain of the1-amino-2-thiol group of native [B] is the only unoxidised sulfhydryl side chain present in native [B].

12. The process according to any one of claims 1, 2, 4, 5, 7 or 10, wherein native [B] comprises additional free thiols which are removed by mutation such that the unoxidised sulfhydryl side chain of the 1-amino-2-thiol group of native [B] becomes the only unoxidised sulfhydryl side chain present.

13. The process according to any one of claims 1, 2, 4, 5, 7 or 10-12, wherein native [B] has a terminal cysteine moiety comprising the terminal 1-amino-2-thiol group.

14. The process according to any one of the preceding claims, wherein the thiol reactive function (TRF) group is a thiol, an alkyl halide, a 2-pyridyl disulfide, a 4-pyridyl disulfide or a Michael acceptor (eg. a vinyl sulfone or a maleimide).

15. The process according to any one of the preceding claims, wherein the thioester group or functional group of native [A] is a benzyl, ethyl or 2-aminoethyl group.

16. The process according to any one of the preceding claims, wherein native [A] and/or native [B] and/or native [C] comprise any one of the following entities or derivatives thereof: pharmacophore, ligand, small molecule, purification handle/immunochemical tag (eg biotin, a dinitrophenyl compound), fluorescent moiety (eg. Alexafluor dye, Texas Red dye), solubilising agent, chelating ligand, chelating ligand plus radioimaging agent, therapeutic protein, antibody or fragment thereof, peptide, peptidomimetic or macroscopic particle (eg. a bead).

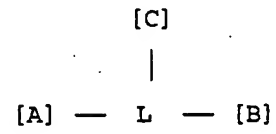
17. The process according to any one of the preceding claims, wherein native [A] and/or native [B] and/or native [C] comprise an entity from a combinatorial array of chemicals.

18. The process according to claim 17, wherein said combinatorial array is a peptide library.

19. A compound obtainable by the process of any one of claims 1-18.

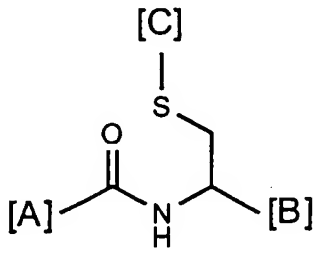
20. A compound obtained by the process of any one of claims 1-18.

21. A compound having the general formula (I):



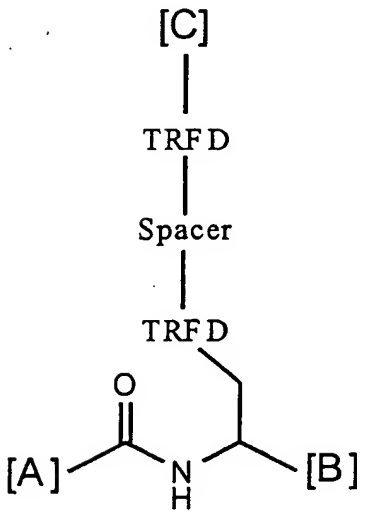
(I)

wherein [A], [B] and [C] are derived from chemical entities native [A], native [B] and native [C], respectively, and are covalently linked by linker group L, the compound of formula (I) having a chemical structure shown in formula (VI):



(VI)

or formula (XI):



(XI)

where TRFD is a thiol reactive function derivative and the Spacer is a linking group,

and wherein each of [A], [B] and [C] comprises any one of the following entities or derivatives thereof: pharmacophore, ligand, small molecule, purification handle/immunochemical tag (eg biotin, a dinitrophenyl compound), fluorescent moiety (eg. Alexafluor dye, Texas Red dye), solubilising agent, chelating ligand, chelating ligand plus radioimaging agent, therapeutic protein, antibody or fragment thereof, peptide, peptidomimetic or macroscopic particle (eg. a bead).

22. Use of the compound according to any one of claims 19-21 in a binding screen for detecting binding to a target entity.

23. The use according to claim 22, wherein the binding screen selects compounds which have increased or decreased binding to a target entity.

24. Use of a compound formed by native chemical ligation in a process for the production of the compound according to any one of claims 19-21.

25. The process, compound or use according to any one of claims 1-24, wherein native [C] has the structure of the second intermediate compound as defined in claim 1 step (iii), the structure of the intermediate compound of formula (II) as defined in claim 2 or 3, the structure of the third intermediate structure as defined in step (v) of claim 4 or claim 7, the structure of the second intermediate compound of formula (IX) as defined in claim 5, or the structure of the second intermediate compound of formula (XII) as defined in claim 6, the compound being a symmetrical homodimer.

26. The process, compound or use according to any one of claims 1-24, wherein the compound is a tetramer.

27. The process according to either one of claim 25 or claim 26, wherein the compound is formed by spontaneous air oxidation between thiol groups or thiol reactive function (TRF) groups.

28. The process, compound or use according to any one of claims 1-27, wherein native [A] and/or native [B] and/or native [C] comprises a myristoyl/electrostatic switch peptide (MSWP) or derivative thereof, for example MSWP2278 (SEQ. ID NO: 2).

29. The process, compound or use according to claim 28, wherein native [A] comprises EGFP, native [B] comprises MSWP2278 (SEQ. ID NO: 2) and native [C] comprises Texas Red Dye.

30. The process, compound or use according to claim 28, wherein native [A] comprises EGFP, native [B] comprises MSWP2278 (SEQ.

ID NO: 2), and native [C] comprises an antibiotic (for example vancomycin).

31. The process, compound or use according to claim 28, wherein native [A] comprises Hirulog anticoagulant peptide, native [B] comprises MSWP2278 (SEQ. ID NO: 2), and native [C] comprises SCR1-3.

Figure 1:

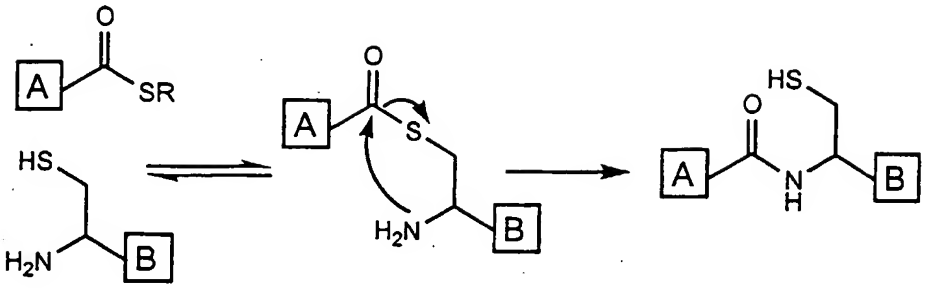


Figure 2:

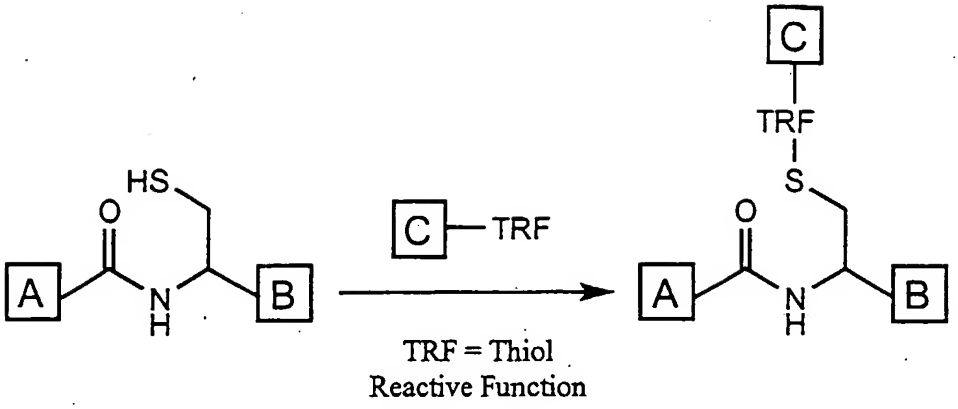


Figure 3:

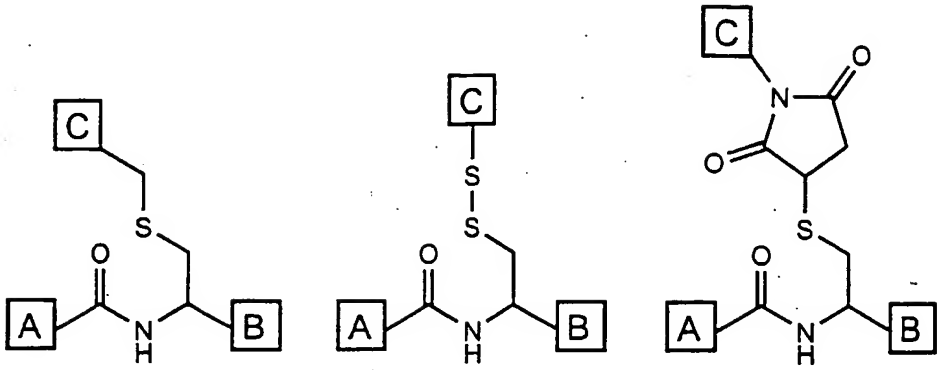


Figure 4:

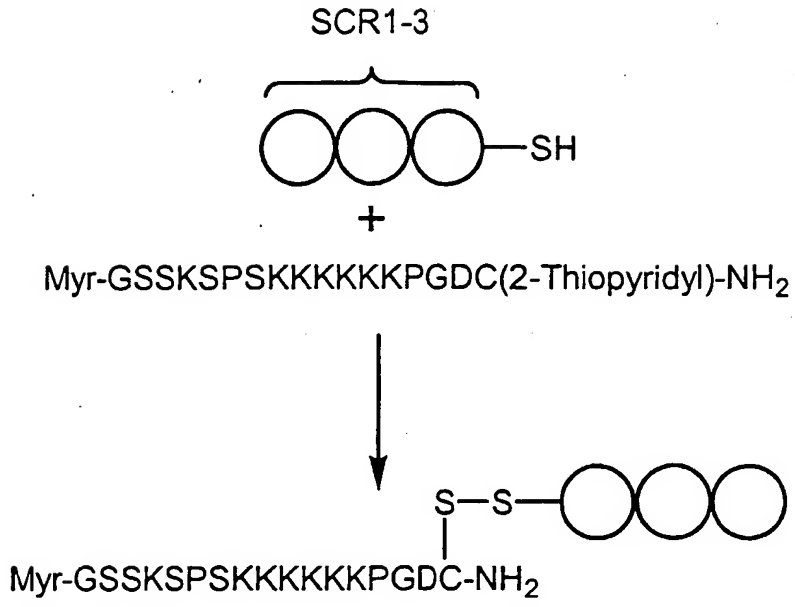

SCR1-3
SH
+
Myr-GSSKSPSKKKKKKPGDC(2-Thiopyridyl)-NH2
S—S
Myr-GSSKSPSKKKKKKPGDC-NH2

Figure 5:

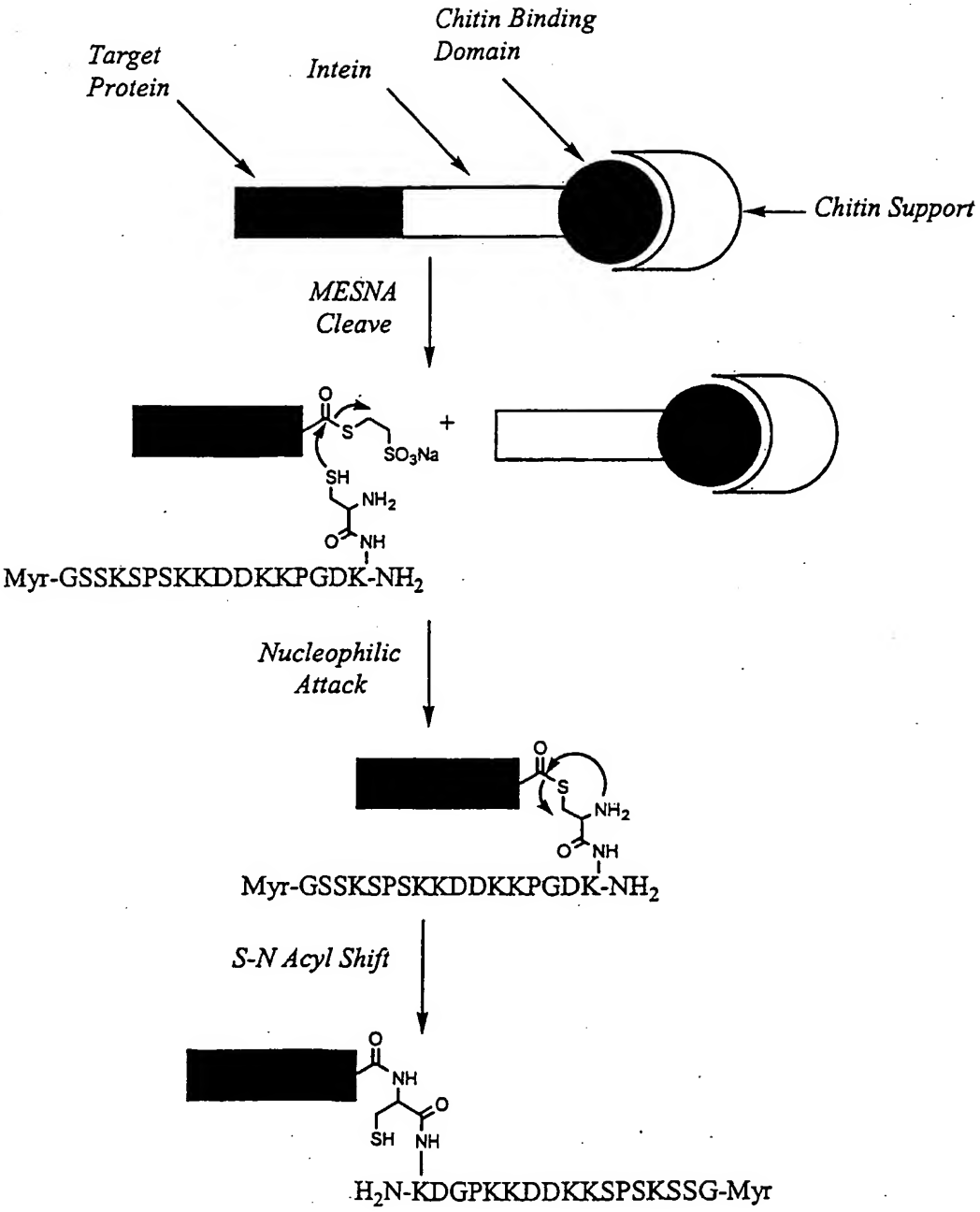

Target Protein
Intein
Chitin Binding Domain
Chitin Support
MESNA Cleave
Myr-GSSKSPSKKDDKKPGDK-NH2
Nucleophilic Attack
Myr-GSSKSPSKKDDKKPGDK-NH2
S-N Acyl Shift
H2N-KDGPKKDDKKSPSKSSG-Myr

Figure 6:

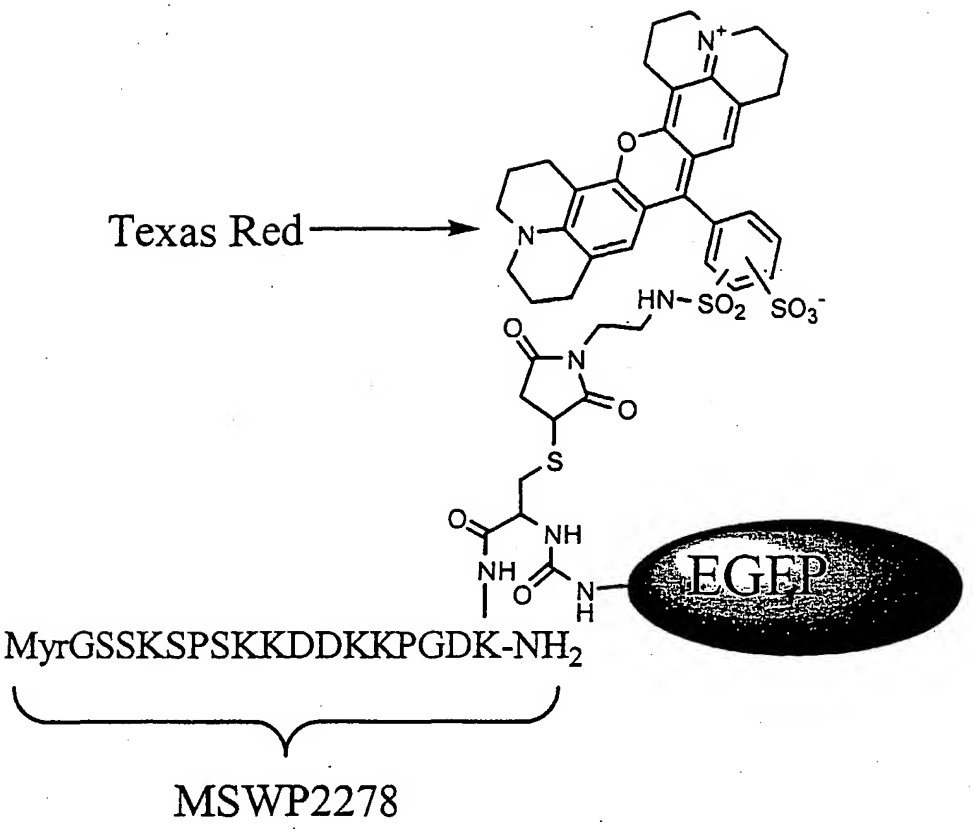


Figure 7:

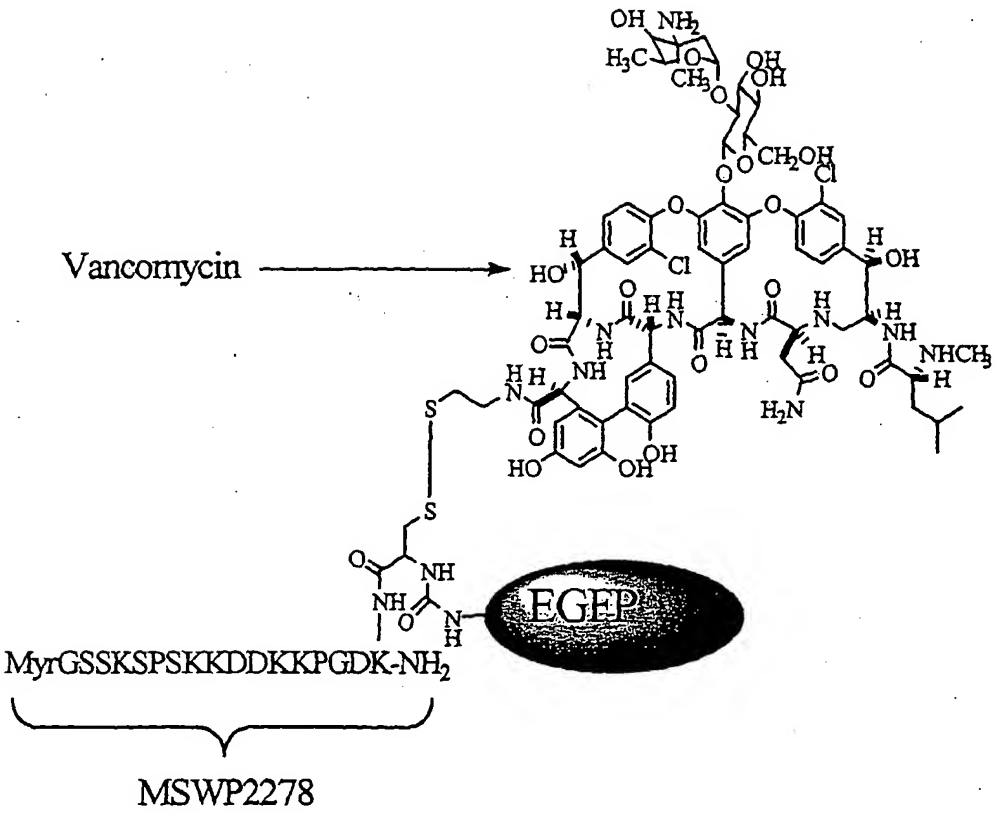


Figure 8:

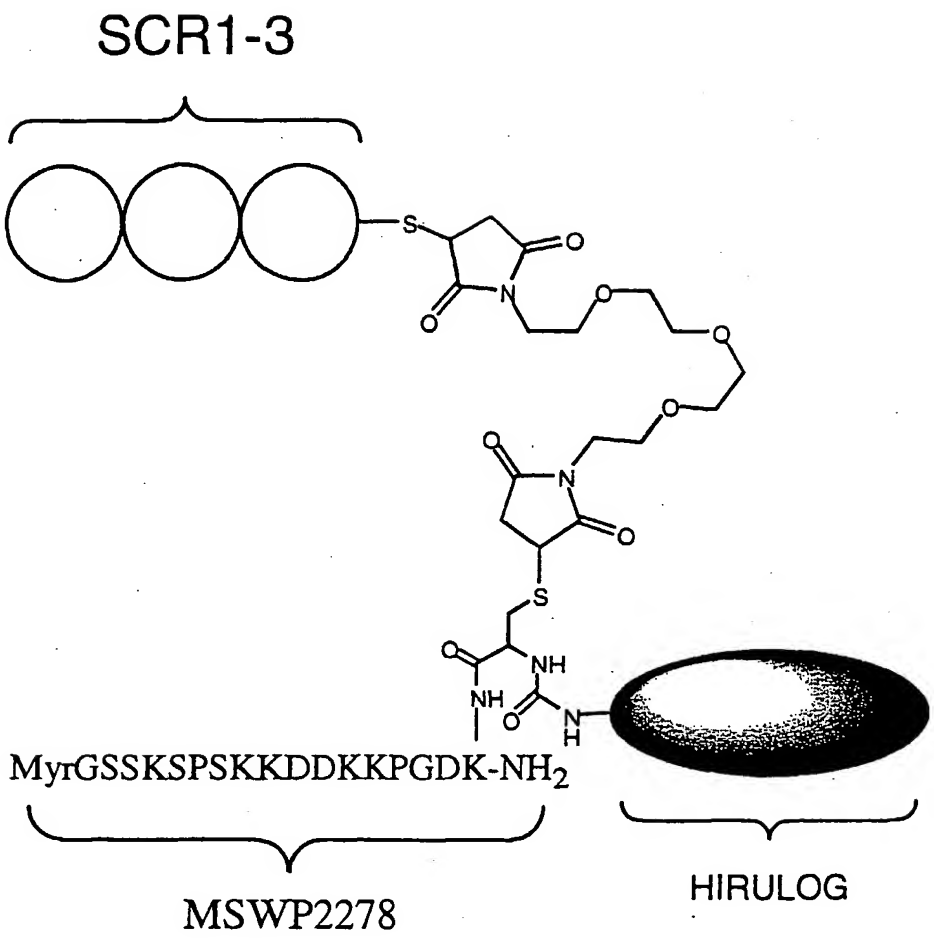
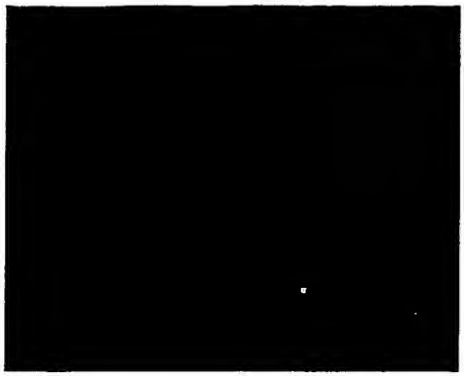


Figure 9:

A



B